

Consolidating CCDs from multiple data sources: A modular approach

Authors

Masoud Hosseini, MSc (Corresponding Author)

Indiana University, School of Informatics and Computing, Department of BioHealth Informatics

Regenstrief Institute, Inc.

Walker Plaza (WK)

719 Indiana Avenue, WK 117

Indianapolis, IN 46202, USA

Email: hosseini@uimail.iu.edu

Jonathan Meade, MSc, Jamie Schnitzius

CreateIT, Inc., Richmond, Indiana, USA

Brian E. Dixon, MPA, PhD

Richard M. Fairbanks School of Public Health at IUPUI

Regenstrief Institute, Inc.

Center for Health Information and Communication, Department of Veterans Affairs, Veterans Health Administration, Health Services Research and Development Service CIN 13-416, Richard L. Roudebush VA Medical Center

Keywords:

De-duplication, Consolidation, Continuity of Care Document (CCD), Health Information Exchange (HIE), Health Level Seven (HL7), Meaningful use, Medical informatics

This is the author's manuscript of the article published in final edited form as:

Hosseini, M., Meade, J., Schnitzius, J., & Dixon, B. E. (2016). Consolidating CCDs from multiple data sources: a modular approach. *Journal of the American Medical Informatics Association*, 23(2), 317–323. <https://doi.org/10.1093/jamia/ocv084>

Abstract

Background: Health care providers sometimes receive multiple Continuity of Care Documents (CCDs) for a single patient encompassing the patient's various encounters and medical history recorded in different information systems. It is cumbersome for providers to explore different pages of CCDs to find specific data which can be duplicated or even conflicted. This study describes initial steps towards a modular system that integrates and de-duplicates multiple CCDs into one consolidated document for viewing or processing patient-level data.

Materials and Methods: We developed a prototype system to consolidate and de-duplicate CCDs. The system is engineered to be scalable, extensible and open source. Using a corpus of 150 de-identified CCDs synthetically generated from a single data source with a common vocabulary to represent 50 unique patients, we tested the system's performance and output. Performance was measured based on document throughput and reduction in file size and volume of data. We further compared the output of the system with manual consolidation and de-duplication. Testing across multiple vendor systems or implementations was not performed.

Results: All of the input CCDs were successfully consolidated, and no data were lost. De-duplication significantly reduced the number of entries in different sections (49% in Problems, 60.6% in Medications, and 79% in Allergies) and reduced the size of the documents (57.5%) as well as the number of lines in each document (58%). The system executed at a rate of approximately 0.009 to 0.03 seconds per rule depending on the complexity of the rule.

Discussion and Conclusion: Given increasing adoption and use of Health Information Exchange (HIE) to share data and information across the care continuum, duplication of information is inevitable. A novel system designed to support automated consolidation and de-duplication of information across clinical documents as they are exchanged shows promise. Future work is needed to expand the capabilities of

the system and further test it using heterogeneous vocabularies across multiple HIE scenarios.

BACKGROUND AND SIGNIFICANCE

To make accurate care and treatment decisions, clinicians require complete, timely and relevant information about a patient. Unfortunately clinical information is fragmented across many different organizations and heterogeneous computer systems. Factors such as variations in insurance coverage, reliance on multiple providers, increasing specialty care, and disparate information systems contribute to information fragmentation (1). Fragmentation creates clinical workflow inefficiencies, leads to over-utilization of health care through duplicative testing or imaging, and prevents coordination of care across healthcare systems (2, 3). Health Information Exchange (HIE) can ameliorate fragmentation by facilitating communication between disparate health information systems (4-6).

Independent information systems providing clinical information are usually syntactically or semantically incompatible, making HIE challenging (7). Standards such as Health Level Seven (HL7) support HIE amongst disparate information systems. HL7 is one of the leading standards for exchange of clinical and administrative data among healthcare information systems and is the most widely recognized international standards development organization in the health care domain(8, 9).

The CCD is an electronic document exchange standard developed by HL7 for sharing patient summary information between providers and organizations. The CCD constrains the HL7 Clinical Document Architecture (CDA) and is provided as a template to improve the interoperability between clinical information systems (10). Further, the CCD standard is required for HIE activities under the Meaningful Use criteria, regulations from the CMS that incentivize the adoption and use of interoperable clinical information systems (11).

Given pressure from health reform and market forces, clinical provider organizations and their information system vendors are required to generate and process CCD documents that contain current

and past medical information about a patient. Sometimes a clinical information system or HIE organization must aggregate multiple CCDs received from a variety of sources and generate a consolidated CCD to facilitate delivery of complete information to a clinician. Care providers prefer to review consolidated information that represents a single, comprehensive picture of a patient's medical history and current condition, rather than multiple CCDs that may include duplicate or conflicting information. Providing too many documents or too much information could require providers to review many pages of a medical record to find a proverbial 'needle in the haystack', which could affect the quality of care delivered to patients.

HIE organizations, which facilitate the electronic transfer of clinical information among a group of health care organizations, seek efficient methods for consolidating multiple CCDs for a given patient.

Consolidation of CCDs, however, is challenging and requires processes that can interpret, merge, de-duplicate, and resolve conflicts across complex documents involving complex data types. Given that meaningful use regulations and many HIE organizations are less than three years old, there do not exist a plethora of proven solutions for consolidating CCDs. HIEs need novel, innovative ideas and products to assist them in managing the growing volume of CCDs available for a patient in support of health care reform goals that seek to provide higher quality, less costly health care to a greater number of people.

In this article, we describe a modular approach to the consolidation and de-duplication of CCDs. The approach is designed to support instances of HIE where there is the need to consolidate multiple clinical documents prior to presentation of information to clinicians or consumption into electronic health record (EHR) systems. We further describe an initial prototype system that embodies the approach.

Although the prototype system is limited in its focus on consolidating three sections of a CCD (e.g., allergies, medications, problem list) and cannot resolve disparate information representations, the approach is nonetheless an important first step. The rest of the paper is organized as follows. Section 2

describes the design of the approach, prototype system and the methods used to pilot data acquisition and analysis. The prototype system was evaluated using synthetically generated CCDs from a single HIE with a common local vocabulary and did not consider disparate CCDs from multiple vendor systems with heterogeneous vocabularies. Section 3 is dedicated to the results of the pilot system evaluation and feasibility testing. Finally, Section 4 summarizes our conclusions and proposes future research directions, which include evaluation using real-world documents from multiple vendor systems and expansion to resolve semantic meaning from disparate information representations.

MATERIALS AND METHODS

In order to aggregate multiple CCDs received from different sources and HIEs, we developed a modular solution to combine the CCDs into a single document and resolve the conflict and duplication of multiple documents for each patient.

Study Context

The design and development of our approach originated as a component of the 2013 Hoosier Healthcare Innovation Challenge (HHIC). HHIC is a state-based software development competition that brings together healthcare and technology professionals to provide creative solutions for some of the most challenging problems in healthcare (12). As a part of the 2013 HHIC, the Indiana Health Information Exchange (IHIE) proposed CCD consolidation and de-duplication as a ‘grand challenge’ facing the HIE marketplace. IHIE is one of the largest HIE organizations in the U.S., serving 25,000 physicians and over 10 million patients (13, 14). The HIE includes multiple integrated delivery networks, hospitals, physician practices, laboratories, radiology centers, long term post-acute care facilities, and public health agencies.

In recent years, IHIE observed that multiple HIE partners, including other HIE networks in Indiana, members of the eHealth Exchange national HIE, and private enterprise health systems within its own

network, began implementing CCD interfaces to send and receive information about individual encounters as well as a patient's past medical history (15, 16). A total of 58 facilities within the IHIE network regularly send CCDs within the exchange. The growth in CCD interfaces was driven by meaningful use incentives, which emphasize the CCD as a preferred message type for providing a summary of care as patients' transition in the health system.

In order to meet customer demands, IHIE needed to gather a wide range of CCDs 'just-in-time,' consolidate the information across documents, and provide a single record of summarized medical information to an application or end user. For example, between July 2014 and April 2015, a total of 767,758 CCDs were exchanged for 359,503 unique patients. While many patients only had one CCD sent, the number of CCDs per patient ranged from 1 to 17 with the number of facilities ranging from 1 to 13. Over 12,000 patients had at least 2 CCDs exchanged during the 9 months. Given the variety of CCD documents potentially available for a given patient, IHIE perceived the need to implement a solution to consolidate and de-duplicate CCDs as they are exchanged across its network. To address this problem, the authors formed a team for the HHIC competition and developed an approach which was ultimately accepted by IHIE and selected as a best solution in the 2013 HHIC competition.

System Description

We designed the system to meet three paradigms in computer science: scalability, extensibility, and open source. IHIE desired the solution to provide for evolving needs as the HIE market and health IT standards evolved. This required the solution to be scalable, or expandable to greater numbers of interfaces and volumes of CCDs over time, and configurable as IHIE's customer needs or CDA document types evolved over time. Currently IHIE and its participants are exchanging the HITSP/C32 - Summary Documents Using HL7 Continuity of Care Document (17), a variety of CDA document as required for Stage 1 of the U.S. Centers for Medicare and Medicaid Services (CMS) "meaningful use" program (18).

However, this is expected to evolve towards the consolidated CDA (C-CDA) with later stages of meaningful use adoption by IHIE and its participants. Furthermore, Regenstrief, IHIE and their partners believe firmly in an open source philosophy or that the system should be available and extensible by not only the original developers but the entire HIE community. Therefore, we chose a modular approach implemented on an open source stack.

The system architecture (Figure 1) consists of three primary components: 1) a CCD Consolidation Engine that receives multiple CCD documents through an application programming interface (API) and executes a series of rules to consolidate and de-duplicate data; 2) an Audit component that identifies and logs system events; and 3) a configuration component that enables implementers to configure the rules executed by the system. Currently, the system is designed with a RESTful API and a simple web application to enable users to configure their desired rule set but this can easily be swapped out to meet the needs of the application calling the system. Multiple CCDs are fed into the Consolidation Engine and the system returns a merged CCD after executing variety of rules defined and configured in Rules engine. During the execution of processes, audit information is captured into a NoSQL database.

<Insert Figure-1 approximately here>

CCD Consolidation Engine

The heart of the solution is the CCD Consolidation Engine. The engine is designed to execute a set of rules to consolidate and de-duplicate information across a set of CCDs into a single CCD. Four classes of rules were developed:

Pre-Format Rules examine incoming CCDs and inspect them for quality. Experience has shown that electronic clinical messages can vary in their interpretation of HL7 standards (19), leading to imperfect messages that contain important data. Pre-Format rules can normalize incoming messages and enable imperfect messages to be processed (20). The Pre-Format rules can execute functions such as validating

CCDs (e.g., applying criteria established by HITSP and HL7) (21-23), dropping invalid CCDs (if this is desired), or strip unwanted sections of the CCD out.

Primary Rule ensures that the program returns something to the calling application. The idea is that one record in the list of CCDs is chosen as the “master CCD.” Selection of the master CCD is based on a simple rule: choose the one with the highest number of entries (e.g., the biggest). The other CCDs are merged into the master CCD. If the merger fails for any reason, then the master CCD along with the non-merged CCDs will be returned. This rule prevents the system from failing to return anything and potentially leaving a clinician without any information to view, something that can lead to user frustration and abandonment of health IT systems.

De-duplication Rules merge, de-duplicate and handle conflicting information. The initial prototype contains nine main classes of rules. Some classes focus on examining and consolidating specific sections: e.g., Medications, Problems, Vital Signs, and Immunization. Other classes focus on resolving conflicts for certain attributes of CCD entries: e.g., Assigned Authors, Confidentiality. Each class is configurable and expandable to meet local needs. For example, in Medication class, functions extract data from the entry level of each CCD. The extracted information includes Medication date, Medication Status Code (e.g., active), administrationUnitCode (e.g., Tablets), manufacturedMaterial (e.g., Raloxifene 60 MG Oral Tablet), Medication Code (e.g., 312768), Medication CodeSystem (e.g., 2.16.840.1.113883.6.88), Medication OriginalText (e.g., Evista), Medication Name (e.g., Evista), and healthcare provider information. Then through firing multiple rules (usually comparison conditions), duplicated entries are discarded and the remaining items are merged into master CCD. To enable comparison, the class requires a minimum of three fields (code, code system, and healthcare provider name). If an entry does not possess at least these fields, then it is automatically retained or merged into the master CCD.

Post-Format Rules check the consolidated document for internal validity and ensure the format of the

document conforms to applicable HL7 CDA and CCD constraints.

<Insert Figure-2 approximately here>

Audit

After each rule in the Consolidation Engine is executed, the engine sends the list of input CCDs, the merged CCD, discarded data elements, and information on the executed rule to the audit system. The audit system stores this information in a NoSQL database, such as MongoDB. The NoSQL database is used because of the large volume of data being sent by the audit system. NoSQL databases are designed to deal with large amounts of data for long term storage. Figure 2 presents a sequence diagram for consolidation of three CCDs through our application including feeding, merging and auditing processes.

User Configuration

An application layer was developed on top of the Consolidation Engine to enable user configuration. A simple demo web application (Figure 3) presents how user can select a set of rules to be fired against CCDs.

<Insert Figure-3 approximately here>

Study Design

We evaluated the system using a corpus of 150 CCDs representing 50 unique patients sampled from IHIE's over 10 million patients (14, 24). Although a variety of rules have been developed in the Consolidation Engine, we focused the evaluation on three important sections of the CCD: Medications, Problems List, and Allergies. These sections contain key information necessary to perform important clinical tasks, such as medication reconciliation, drug-drug interaction checking, and drug-allergy checking.

The CCDs were input into the system for consolidation and de-duplication in batches of three

documents where the three documents simulated disparate institutional data for the same patient. The system output was a single, consolidated CCD. We examined system performance as well as the difference between the input documents and the output document. Our analysis focused on whether the system correctly identified and de-duplicated the same information spread across three disparate CCDs.

The sample size is sufficient for detecting a difference in the size or number of entries for 50 (N) patients with a moderate effect size (0.4) at 80% power. For a corpus of 150 CCDs (3N), the largest decrease possible when the documents are exactly the same is 67%; but if there are duplicates within each report, the largest decrease would be $\geq 67\%$. While one could choose to evaluate a range (2N, 3N, 4N, 5N) of options with respect to the number of CCDs, 3N has some basis in reality: an emergency encounter with resulting hospital stay followed by primary care follow-up; or a primary care visit followed by specialty visit followed by primary care follow-up. In these scenarios, a CCD would be generated when there is a transfer of care. While a larger sample of patients would be necessary to detect smaller effect sizes, we anticipated at least moderate effect sizes given our methods for generating the synthetic CCDs.

Methods for Data Acquisition

To generate the corpus of CCDs for testing, we first used the production instance of the Regenstrief CCD generator – a web service that generates a CCD for a given patient – to create de-identified CCDs for 50 unique patients selected from a population of 12 million. We randomly selected patients with a minimum of seven active problems, five active medications, and three documented allergies. We then twice cloned the CCDs to create multiple input documents for each patient. Next we manipulated the cloned CCDs to simulate disparate clinical information drawn from various EHR systems. An algorithm manipulated the medication, problem list, and allergy sections of the CCDs, randomly deleting existing entries and/or adding new entries to the documents.

After the clones were manipulated, the corpus contained 150 unique, de-identified documents for 50

unique patients. Manipulated CCDs were fed into the CCD consolidation service, in sets of 3, and merged CCDs were generated for each patient. When input into the system for testing, the system could not distinguish between the original or manipulated clone CCDs.

Methods for Data Analysis

System performance was determined based on the total time necessary to process the corpus of 150 input CCDs. We examined the average time to process a CCD as well as the minimum, maximum, and average file sizes. To examine accuracy, we manually de-duplicated the CCDs and compared the manually merged document with the consolidated CCD generated by the system.

RESULTS

All 150 input CCDs were successfully consolidated into 50 output CCDs. Table-1 summarizes the number of entries within each section of the CCDs before and after consolidation. The consolidation process significantly reduced the number of entries due to duplicate values in each section that the engine identified and removed during the consolidation process. The consolidation process also reduced the size of the documents, as well as the number of lines in each document. Manual review of the results confirm that the Consolidation Engine correctly identified all (100%) of the duplicate entries.

	A: Number of Entries in Input CCDs (N=150)	B: Number of Entries in Consolidated CCDs (N=50)	Percentage Decrease from A to B
Problems Section	7,511	3830	49%
Medication Section	4,822	1900	60.6%
Allergy Section	1,309	275	79%
File size on disc (KB)	92,205	39,199	57.5%
Number of lines in CCDs	1,636,385	687,192	58%

Table 1: Comparison of CCD properties before and after consolidation

The current prototype is executing consolidation and de-duplication rules against the test dataset at a rate of approximately .009 to .03 seconds per rule depending on the complexity of the rule. The overall de-duplication time for 150 CCDs is 169 seconds which covers the time for firing rules on our focused sections (Problems, Medications, and Alert) and other rules such as Primary Merge Rule, Vital Signs, and Confidentiality.

The audit write method takes longer than the actual consolidation but the auditing methods can be assigned to a separate processor thread to run, which could improve performance. According to the Audit logs, the largest file size among the 150 CCDs is 6912 KB and the smallest one is 90 KB. The overall file size of 150 CCDs is 90.04 MB. For this size of data, the number of de-duplication occurred in Problems section is 3686 and this number is 2923 for Medications and 1035 for Alerts section. The overall number of entries in Problems section is 7511, in Medications 4822, and in Alerts section is 1309 (Table 1).

DISCUSSION

Consolidating and de-duplicating information in clinical summary documents is a requirement of HIE systems, because data are fragmented and often duplicated across providers. In this study, we performed preliminary evaluation of a CCD Consolidation Engine designed to merge and de-duplicate CCDs as they are exchanged for consumption into an EHR or immediate presentation to a clinical user. Preliminary results are favorable, showing that with simulated data the consolidation engine was able to identify all (100%) of the duplicated entries for a limited set of document sections across groups of CCDs for the same patient.

Decreasing the number of entries (49% in Problems entries, 60.6% in Medications entries, and 79% in Allergies entries) may help providers to read medical information without confronting duplicate entries

which may cause frustration or safety concerns. For example, when reconciling medication data to determine potential changes in a patient's drug regimen, duplicate entries may require additional time to review the list leading to clinician frustration or they may be hard to detect leading to unintended consequences. Such a scenario would only exacerbate clinician fatigue and usability challenges presented by HIE systems (25, 26). Furthermore, consolidation significantly reduces the size of CCDs, which has an impact on the time required to process and present information to users in real-world systems. Even if there is no duplication among a group of CCDs, consolidation of documents reduces the overall size of the information because, at minimum, patient demographics are repeated across CCDs. This may speed processing of the information and potentially aid in the display of the information to end users. Therefore the consolidation may be a useful component of operational HIE systems, or implemented within a hospital or health system to process incoming or outgoing CCDs for a given patient before consuming the potentially duplicative information or rendering it to a clinical end user. System performance is adequate for many use cases but insufficient for others. Total time to process 150 CCDs is 169 seconds, just over one second per document. This is likely reasonable in an automated use case where input CCDs are considered for consumption into an EHR system, say, during an overnight batch process. However, this performance may not be adequate in a scenario where a clinician is waiting to view documents. In the Department of Veterans Affairs, providers can access multiple CCDs from non-VA providers through an initiative called the Virtual Lifetime Electronic Record (VLER) (15). Clinicians initiate requests for external information by opening the VA's VistaWeb application, the component within its enterprise EHR that enables access to fragmented information captured by other VA facilities, as well as the Department of Defense. Currently the end user receives a potentially long list of CCDs for the patient from various VLER partners – usually private HIE organizations but could be another federal agency. Applying the CCD Consolidation Engine might be useful to reduce the number of

documents to be rendered to the end user. However, if the process takes more than one second to complete, the end user may become frustrated. Past research at Regenstrief and others has demonstrated that clinicians expect sub-second performance from clinical information systems (27). Furthermore, the results from this pilot using similarly structured CCDs may not hold up in a larger evaluation, therefore additional testing and evolution will be necessary to make the system robust under various conditions and contexts.

Although the CCDs generated by the Regenstrief Institute are conformant with HL7 standards, we have observed that many real-world CCDs being exchanged within HIEs as well as health systems are not always conformant. The Indiana HIE supported by Regenstrief utilizes a wide range of pre-processors to fix incoming messages from over 1000 source systems (20). These pre-processors help to add missing document structure components or move data from say a <text> block to a <observation> block if one does not exist. The CCD Consolidation Engine assumes that the received documents for consolidation are valid. Yet it is possible that non-conformant CCDs may get inserted into system at some point. So, instead of ignoring or deleting invalid documents, we designed our solution to be lenient in what it accepts with the intention of creating a robust solution for the heterogeneity that exists in real-world clinical messages. For instance, when there is no valid template ID for the Medications section in a CCD and the values for “code” or “code system” are not valid, the application tries to parse the title of that section and if there is a word “Medications” in title, it assumes that section as Medications section. However, if there is no such useful information, the application skips that section.

Limitations and Future Development

Although the results of this evaluation are promising, the system has several limitations. First, because of the competition time constraints, we used synthetic data output from a single CCD generation engine. Second, while the system can detect conflicting information across documents there are few rules for

how to adjudicate discrepancies. Third, the engine is not yet sophisticated enough to detect semantic overlap when varying code systems are present or free text is used.

This evaluation focused on synthetic data output from a single CCD generator, one developed by Regenstrief and used by IHIE. Although the 50 base CCDs were derived from real patients, the resulting sample was nonetheless small and derived from a single source. Future work will evaluate the approach and system using real-world CCDs exchanged as part of meaningful use from multiple sources. We will test the system using multiple CCDs exchanged among 60 facilities representing five health systems and more than 10 independent community hospitals.

The engine currently possesses several rules to detect conflicting information. For instance, one CCD may show that a patient is allergic to penicillin but the other CCD reports the patient has no known allergies. Another example is where one CCD documents a patient response to a question about Cigarette-Use as “Never Smoked Cigarette” but in another document the value is “Smoker.” Currently the consolidation engine detects these conflicts, but the system is limited in what it does with such information. Presenting these conflicts after detection is challenging because there does not exist a standard representation of this kind of information in CCD documents. In addition, while the current engine contains several such rules, there are many more varieties of conflicts. Future work will be necessary to expand the list of conflicts examined by the system. Moreover, once discovered there is no clear logic for adjudicating these kinds of conflicts. When to keep such conflicts or whether it is possible to make *a priori* determinations about which values might take precedence over others (e.g., always keep presumed allergies even if one document explicitly reports no known allergies) need to be addressed in future work and in a systematic way.

While the engine performed well detecting duplicates, the kind of duplicate entries it can detect is limited to cases where the same semantic encoding is used in multiple CCDs. The engine cannot

currently detect semantic relationships between documents where different coding systems are used (e.g., SNOMED vs. ICD) or between encoded entries and free text entries (e.g., SNOMED vs. “Type 2 diabetes”). Future work will be necessary to enhance the engine to detect true semantic relationships we know exist in the real-world (28). Approaches that leverage the UMLS Metathesaurus may be explored to take advantage of already cross-linked ontologies. Natural language processing (NLP) approaches may also be necessary to link non-encoded observations in one CCD with encoded entries in other CCDs.

CONCLUSION

The use and exchange of CCDs are expected to increase during the next phase of meaningful use. It is helpful that providers can access patients’ medical records from multiple locations through HIE, however, duplication and conflict of information is inevitable. The potential of having a document consolidation engine as a service, is a promising solution for HIEs to provide information that is easier to consume for clinical providers and more findable because of integration of homogenous information in one section. Our effort in this study is to present a novel prototype for de-duplication and consolidation of CCDs, which can be considered a first step towards improving interoperability among information systems and information access among providers.

FUNDING STATEMENT

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

COMPETING INTERESTS STATEMENTS

The authors have no competing interests to declare.

CONTRIBUTORSHIP STATEMENT

M. Hosseini and BE Dixon conceived of the approach and study design. M. Hosseini, J. Mead, and J.

Schnitzius contributed to the development of the system and rules engine. M. Hosseini conducted the system testing and evaluation, including the extraction of data and generation of the synthetic documents used to assess feasibility. BE Dixon provided guidance on the evaluation and interpretation of the results. M. Hosseini prepared the draft version of the article. BE Dixon supervised the study and provided critical review and edits to the article. All authors contributed to refinement of the study protocol and approved the final manuscript.

References:

- 1 Vest JR, Gamm LD. Health information exchange: persistent challenges and new strategies. Journal of the American Medical Informatics Association. 2010;**17**(3):288-94.
- 2 Schoen C, Osborn R, Squires D, et al. A Survey Of Primary Care Doctors In Ten Countries Shows Progress In Use Of Health Information Technology, Less In Other Areas. Health affairs (Project Hope). 2012 Nov 15.
- 3 Frisse ME, Johnson KB, Nian H, et al. The financial impact of health information exchange on emergency department care. J Am Med Inform Assoc. 2012 May-Jun;**19**(3):328-33.
- 4 Brailer DJ. Interoperability: the key to the future health care system. Health Affairs-Millwood Va Then Bethesda MA. 2005;**24**:W5.
- 5 Acker B, Birnbaum C, Braden J, et al. HIM principles in health information exchange. Journal of AHIMA/American Health Information Management Association. 2007;**78**:69-74.
- 6 Hripcsak G, Kaushal R, Johnson KB, et al. The United Hospital Fund meeting on evaluating health information exchange. Journal of biomedical informatics. 2007;**40**(6):S3-S10.
- 7 Dixon BE, Vreeman DJ, Grannis SJ. The long road to semantic interoperability in support of public health: experiences from two states. J Biomed Inform. 2014 Jun;**49**:3-8.
- 8 Hosseini M, Ahmadi M, Dixon BE. A Service Oriented Architecture Approach to Achieve Interoperability between Immunization Information Systems in Iran. AMIA Annu Symp Proc; 2014: American Medical Informatics Association; 2014.
- 9 2014 Edition Release 2 Electronic Health Record (EHR) Certification Criteria and the ONC HIT Certification Program; Regulatory Flexibilities, Improvements, and Enhanced Health Information Exchange. The U.S. Department of Health and Human Services; 2014.
- 10 Dolin RH, Giannone G, Schadow G. Enabling joint commission medication reconciliation objectives

with the HL7 / ASTM Continuity of Care Document standard. AMIA Annu Symp Proc. 2007:186-90.

11 Medicare and Medicaid programs; electronic health record incentive program--stage 2. Final rule.

Federal register. 2012 Sep 4;**77**(171):53967-4162.

12 Bodenheimer T, Wagner EH, Grumbach K. Improving primary care for patients with chronic illness.

Jama. 2002 Oct 9;**288**(14):1775-9.

13 Biondich PG, Dixon BE, Duke J, et al. Regenstrief Medical Informatics: Experiences with Clinical

Decision Support Systems. In: Greenes RA, editor. Clinical Decision Support: The Road to Broad

Adoption. 2 ed. Burlington, MA: Elsevier, Inc.; 2014. p. 165-87.

14 Biondich PG, Grannis SJ. The Indiana network for patient care: an integrated clinical information

system informed by over thirty years of experience. J Public Health Manag Pract. 2004 Nov;**Suppl**:S81-6.

15 Byrne CM, Mercincavage LM, Bouhaddou O, et al. The Department of Veterans Affairs' (VA)

implementation of the Virtual Lifetime Electronic Record (VLER): findings and lessons learned from

Health Information Exchange at 12 sites. Int J Med Inform. 2014 Aug;**83**(8):537-47.

16 Dixon BE, Colvard C, Tierney WM. Identifying health facilities outside the enterprise: challenges and

strategies for supporting health reform and meaningful use. Inform Health Soc Care. 2014 Jun 24;1-15.

17 Nielsen PE, Thomson BA, Jackson RB, Kosman K, Kiley KC. Standard obstetric record charting system:

evaluation of a new electronic medical record. Obstet Gynecol. 2000 Dec;**96**(6):1003-8.

18 Health information technology: initial set of standards, implementation specifications, and

certification criteria for electronic health record technology. Final rule. Fed Regist. 2010 Jul

28;**75**(144):44589-654.

19 Barnes M. Lessons learned from the implementation of clinical messaging systems. AMIA Annu Symp

Proc. 2007:36-40.

20 Zafar A, Dixon BE. Pulling back the covers: technical lessons of a real-world health information

exchange. Studies in health technology and informatics. 2007;**129**(Pt 1):488-92.

21 HITSP Summary Documents Using HL7 Continuity of Care Document (CCD) Component. Healthcare Information Technology Standards Panel (HITSP); 2009.

22 HITSP Clinical Document and Message Terminology Component. Healthcare Information Technology Standards Panel (HITSP); 2010.

23 HITSP CDA Content Modules Component. Healthcare Information Technology Standards Panel (HITSP); 2009.

24 McDonald CJ, Overhage JM, Barnes M, et al. The Indiana network for patient care: a working local health information infrastructure. An example of a working infrastructure collaboration that links data from five health systems and hundreds of millions of entries. Health affairs (Project Hope). 2005 Sep-Oct;**24**(5):1214-20.

25 Gadd CS, Ho YX, Cala CM, et al. User perspectives on the usability of a regional health information exchange. J Am Med Inform Assoc. 2011 Sep-Oct;**18**(5):711-6.

26 Dixon BE, Simonaitis L, Goldberg HS, et al. A pilot study of distributed knowledge management and clinical decision support in the cloud. Artif Intell Med. 2013 Sep;**59**(1):45-53.

27 Paterno MD, Goldberg HS, Simonaitis L, et al. Using a Service Oriented Architecture Approach to Clinical Decision Support: Performance Results from Two CDS Consortium Demonstrations. AMIA Annu Symp Proc. 2012;**2012**:690-8.

28 Lin MC, Vreeman DJ, McDonald CJ, Huff SM. A characterization of local LOINC mapping for laboratory tests in three large institutions. Methods of information in medicine. 2011;**50**(2):105-14.

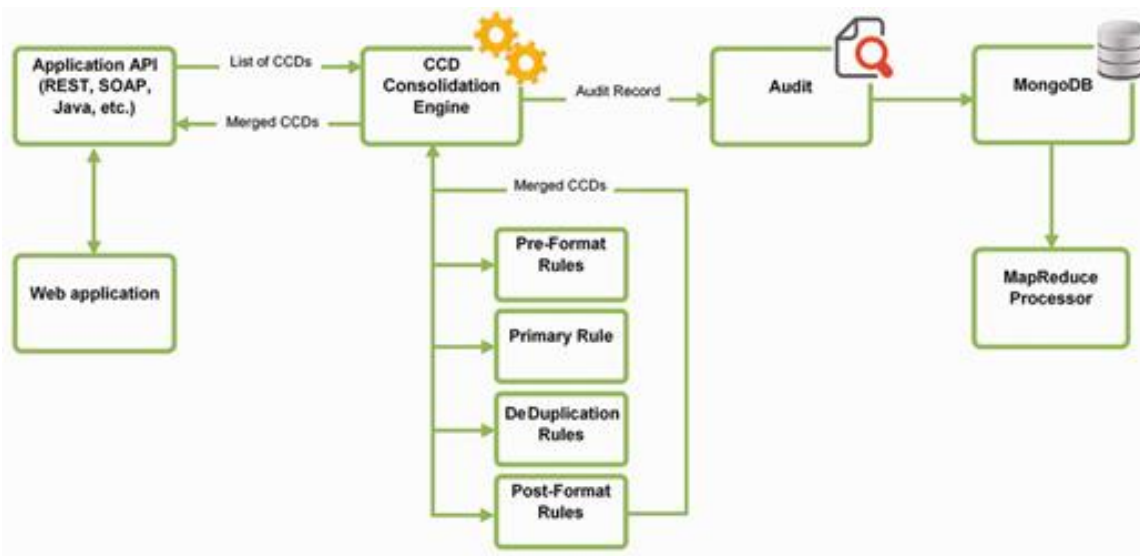
FIGURE LEGENDS

Figure 1. CCD Consolidation Application Model

Figure 2. Depicts the sequence Diagram of consolidation of input CCDs through different modules of the system. This diagram shows that three CCDs are passed through an API to the Rules Engine. Rules are executed and, for every rule, one audit event is recorded (the list of input CCDs, the merged CCD, discarded data elements, and information on the executed rule) into a NOSQL database. Finally, the consolidated CCDs are returned to the calling application or service.

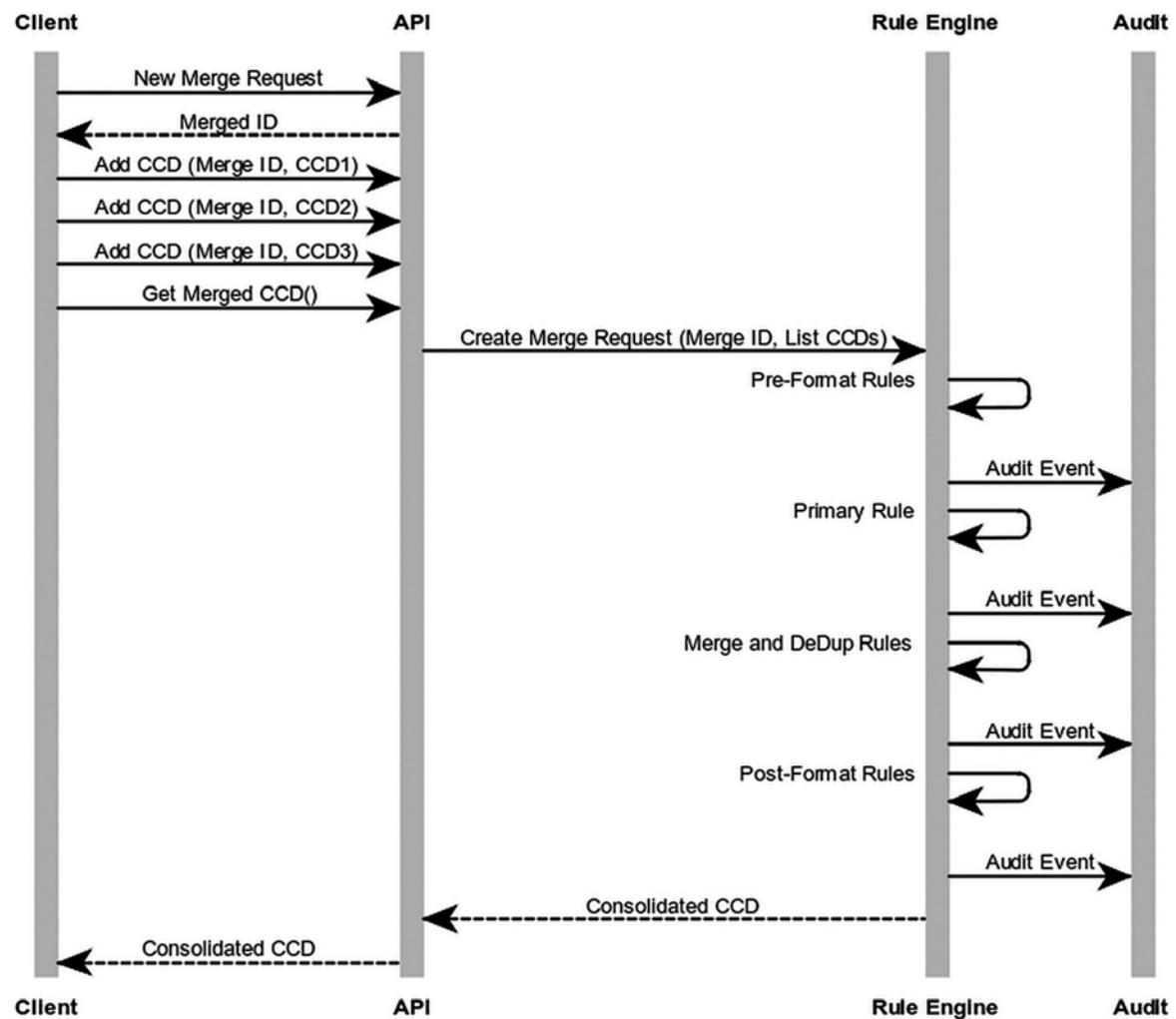
Figure 3. A Web-based user interface that enables selection of the desired rules to be fired against input CCDs. The three input CCDs are loaded into the text boxes at the left side of the figure and the consolidated, de-duplicated CCD in the right side of the figure appears after pushing the test button. The user can change the rule options from the top side of the interface as many time that is needed and the system will generate a new, consolidated CCD based on the selected rules.

Figure 1:



CCD consolidation application model.

Figure 2:



Depicts the sequence diagram of consolidation of input CCDs through different modules of the system. This diagram shows that three CCDs are passed through an API to the Rules Engine. Rules are executed and, for every rule, one audit event is recorded (the list of input CCDs, the merged CCD, discarded data elements, and information on the executed rule) into a NoSQL database. Finally, the consolidated CCDs are returned to the calling application or service.

Figure 3:

Pre-format Rules	Primary Rule	Merge and De-Dup Rules	Post-Format Rules
<input checked="" type="checkbox"/> RULE: Strip Text	RULE: Latest Valid CCD <input checked="" type="checkbox"/>	<input checked="" type="checkbox"/> RULE: Alerts Section <input checked="" type="checkbox"/> RULE: Medications Sections <input checked="" type="checkbox"/> RULE: Problems Section <input checked="" type="checkbox"/> RULE: Social History - Alcohol Use <input checked="" type="checkbox"/> RULE: Social History - Smoking	<input checked="" type="checkbox"/> RULE: Insert Formatted Text

CCD 1

```
<ClinicalDocument xmlns="urn:hl7-org:v3" xmlns:voc="urn:hl7-org:v3/voc" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <!--
  Header section
  -->
  <typeId root="2.16.840.1.113883.1.3"
  extension="POCD_HD000040" />
  </ClinicalDocument>
```

CCD 2

```
<ClinicalDocument xmlns="urn:hl7-org:v3" xmlns:voc="urn:hl7-org:v3/voc" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <!--
  Header section
  -->
  <typeId root="2.16.840.1.113883.1.3"
  extension="POCD_HD000040" />
  </ClinicalDocument>
```

CCD 3

```
<ClinicalDocument xmlns="urn:hl7-org:v3" xmlns:voc="urn:hl7-org:v3/voc" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <!--
  Header section
  -->
  <typeId root="2.16.840.1.113883.1.3"
  extension="POCD_HD000040" />
  </ClinicalDocument>
```

De-Duped Final CCD

```
<ClinicalDocument xmlns="urn:hl7-org:v3" xmlns:voc="urn:hl7-org:v3/voc" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <!--
  Header section
  -->
  <typeId root="2.16.840.1.113883.1.3"
  extension="POCD_HD000040" />
  <templateId root="2.16.840.1.113883.3.27.1776"
  assigningAuthorityName="CDR/R2" />
  <templateId root="2.16.840.1.113883.10.20.1"
  assigningAuthorityName="CCD" />
  <templateId root="2.16.840.1.113883.3.88.11.32.1"
  assigningAuthorityName="HITSP/C32" />
  <!--CCD v1.0 Template Root-->
  <id root="ede-00161780af43-4acf6c1f-137-0" />
  <code code="34133-9" displayName="Summarization of episode note" codeSystem="2.16.840.1.113883.6.1"
  codeSystemName="LOINC" />
  <title>Editha Tester - CCD/C32</title>
  <effectiveTime value="20130603130015+0400" />
  <confidentialityCode code="N"
  codeSystem="2.16.840.1.113883.5.25" />
  <languageCode code="en-US" />
  <recordTarget>
    <patientRole>
      <id extension="1170-8"
      root="2.16.840.1.113883.3.227.2.1" />
      <addr use="HFP">
        <streetAddressLine>510 East Lasalle
      </streetAddressLine>
    </patientRole>
  </recordTarget>
</ClinicalDocument>
```

A Web-based user interface that enables selection of the desired rules to be fired against input CCDs. The three input CCDs are loaded into the text boxes at the left side of the figure and the consolidated, de-duplicated CCD in the right side of the figure appears after pushing the test button. The user can change the rule options from the top side of the interface as many times as needed and the system will generate a new, consolidated CCD based on the selected rules.